Energy-Efficient Hybrid Duplexing Strategy for Bi-Directional Distributed Antenna Systems

Zhongxiang Wei, *Member, IEEE,* Sumei Sun, *Fellow, IEEE,* Xu Zhu, *Senior Member, IEEE,* Yi Huang, *Senior Member, IEEE,* Jingjing Wang, *Student Member, IEEE*

***Abstract*—We propose a bi-directional distributed antenna (DA) system where DAs are capable of working dynamically in hybrid duplex modes: full duplex (FD), half duplex (HD) and sleep, which enables higher degree of freedom and hence much higher energy efficiency (EE) than DA sole-FD systems, with just marginal loss in spectral efficiency (SE). The proposed system also demonstrates significant EE and SE enhancement over FD co-located antenna systems. Compared to the above two benchmark systems, the proposed system also requires much simpler self-interference cancellation (SIC) design for FD mode due to less cross talks between antennas. A low-complexity EE maximization scheme is proposed for the bi-directional DA system. A channel gain based DA clustering algorithm is first performed to activate/deactivate transmit/receive chains, which highlights the characteristics of DA deployment, and then a distributed hybrid duplexing (Dis-Hyb-Duplexing) algorithm is performed to optimize the downlink beamformer and the uplink transmission power. Various practical aspects are taken into account for system design, such as self-interference at DAs in FD mode, co-channel interference from uplink users to downlink users, and multiuser interference in both uplink and downlink. The effectiveness of the proposed system is verified by simulation results.**

***Index Terms*—Hybrid duplexing, full duplex, half duplex, distributed antennas, energy efficiency**

I. Introduction

Full-duplex (FD) multi-input multi-output (MIMO) communication [1] is regarded as a promising solution to meet the rapidly increasing demand for high data rate. Unlike the conventional half-duplex (HD) MIMO, where uplink and downlink transmission are decoupled into orthogonal time slots or frequencies [2] [3], a base station (BS) equipped with FD MIMO enables bi-directional communication with uplink and downlink users at the same time and frequency band [4] [5]. Due to the FD bi-directional transmission, self-interference is introduced at the BS. By self-interference cancellation (SIC) at the BS, *i.e.*, passive suppression [6], analog cancellation and digital cancellation [7] [8], self-interference can be effectively mitigated.

Spectral efficiency (SE) maximization has been extensively researched to exploit the advantage of FD MIMO systems [5] [9] [10] [11] [12]. The authors of [5] investigated the FD transmission in various networks, such as bi-directional relay and cellular networks. The authors of [9] found that by applying mode switch, power control, interference-aware beamforming or link selection, SE of FD transmission can be improved. In [10] [11], SE of FD MIMO bi-directional systems was investigated, where small cell FD BSs were deployed to communicate with multiple users in downlink and uplink simultaneously. In [12], a multiuser FD MIMO precoding transceiver structure applicable for single-carrier and orthogonal frequency division multiplexing (OFDM) systems was presented to maximize SE. FD MIMO systems with heterogeneous cells were researched by [13] and [14] to improve SE. In [13], FD was applied for small cell BSs while HD was applied for macro cell BSs to relax the coverage reduction. In [14], the performance of massive MIMO wireless backhaul systems was investigated, where small cells adaptively work in FD or HD mode and each macro cell BS serves its small cells by using zero forcing beamforming. On the other hand, SE performance of FD relay systems was investigated in [15] [16]. Relay selection was proposed for a two-way FD relay system in [15], where two ends communicate with each other assisted by multiple two-way relays. A relay-assisted three-node system was investigated by [16], where the relay adaptively switches between FD and HD modes and selects the best links for transmission and reception.

Some fundamental challenges, however, need to be addressed in FD MIMO systems. The first challenge is that both MIMO and FD techniques require much higher power consumption, which is against the green evolution requirement proposed by the future communication systems. This is because a) MIMO increases the number of active transmit/receive chains, leading to much higher power consumption than single-input and single-output (SISO) [2]. b) For FD systems, additional power consumption is incurred by SIC [17] [18]. Therefore, how to ensure high energy efficiency (EE) transmission needs to be considered. However, most EE-oriented research only focuses on HD one-directional systems [19] [20] [21], which may not be directly applied to a FD bi-directional scenario due to the residual self-interference, co-channel interference from uplink users to downlink users and multiuser interference. The second challenge is with the increased number of centralized antennas, SIC circuit design

Manuscript received June 7, 2017; revised November 1, 2017 and January 14, 2018; accepted February 20, 2018. Date of publication March 9, 2018; date of current version June 18, 2018. This work was funded jointly by the University of Liverpool and A*STAR. *Corresponding author: Xu Zhu.*

Zhongxiang Wei, Xu Zhu and Yi Huang are with the department of Electrical Engineering and Electronics at the University of Liverpool, Liverpool, China. Email: {hszwei,xuzhu,yihuang}@liverpool.ac.uk

Sumei Sun is with the Institute for Infocom Research, A*STAR, Singapore. Email: sunsm@i2r.a-star.edu.sg

Jingjing Wang is with the Tsinghua University, Beijing, China. Email: chinaeephd@gmail.com.

in analog/digital domain becomes very complicated. The third challenge is the similar level of high path loss (PL) and correlated small scale fading in co-located MIMO systems.

To this end, centralized FD MIMO can be formulated into a distributed manner. Several advantages can be gained: a) Since the contributions of each distributed antenna (DA) may vary practically due to the location of the users [22], system power consumption can be significantly reduced by only activating those DAs contributing the most. b) With distributed deployment, SIC becomes much easier due to the fact that the self-interference among different DAs can be well controlled in the propagation domain, benefiting from the ideal natural isolation performance. As a result, the circuit design for SIC could be as easy as the SISO case within each FD DA itself. c) The DA system can reduce the large-scale fading impact and obtain blockage-free effect using the multiple DAs distributed geographically [23], helping extend coverage and maintain connectivity of networks. In summary, there is strong evidence showing that, DA system is a remarkable alternative for co-located FD MIMO system. DA systems have been extensively utilized in practice to increase SE and extend coverage. In the USA (*e.g.*, Michigan and Ohio), DA systems have been deployed as an important part of the landline infrastructure, as promulgated by the FCC Pole Attachment Order 11-50 [24]. DA systems have also been considered in 5G systems to form user-centric manner virtual cells [25], where users can find DAs in its vicinity to communicate with. In indoor environment, DA systems can be applied to provide seamless coverage [26], such as commercial use of WiFi. Besides, DA systems have been adopted by the International Union of Railways for high-speed rail communication systems [27] [28]. DA systems have also been investigated in academia. In [29], SE maximization and power minimization were investigated for DA systems. In [30], a fairness-aware downlink scheduling algorithm was proposed for DA systems. However, the previous research for DA systems was based on HD transmission only. In [31], FD cooperative multi-point system was studied in terms of power minimization, where the antennas can be set to either FD mode or sleep mode based on the state-of-art switch on/off technique [32]. Having only FD or sleep mode may not be optimal in some scenarios. For example, if the active antennas are only close to uplink users while are very far from downlink users, these antennas may work in HD (receive) mode to incur lower power consumption and absence of self-interference. Since HD mode is featured by lower power consumption and FD mode is featured by high SE at the cost of high power consumption, the hybrid duplexing technique enables antennas to adaptively switch among FD, HD and sleep modes, making a good trade-off between the SE and power consumption.

In this paper, we propose a bi-directional DA system where DAs are capable of working in hybrid duplex mode: FD, HD and sleep, and investigate EE maximization with dynamic hybrid duplexing. The effectiveness of the proposed system is verified by simulation. Our work is different in the following aspects:

1) Hybrid duplex mode enables higher degree of freedom and hence much higher EE compared to DA FD-sole systems [31], with just marginal loss in SE. The proposed system also demonstrates significant EE and SE enhancement over co-located FD MIMO systems [11].

2) Compared to DA sole-FD and co-located FD MIMO systems, the proposed DA hybrid duplexing system requires much simpler SIC design for FD mode, due to less cross talks between DAs.

3) We investigate EE maximization by considering both bi-directional SE and power consumption. This is different from the existing work on SE maximization alone [9]-[11] or total power consumption minimization alone [31]. Various practical aspects are taken into account for the optimization, such as residual self-interference at DAs in FD mode, co-channel interference from uplink users to downlink users, and multiuser interference in both uplink and downlink, whereas only self-interference was considered in our previous work on one-directional FD relay systems [4] [18]. Also, the power amplifier (PA) dissipated power, circuit power, system fixed power and SIC power are all included in the power consumption model, which is more accurate than the existing models formulated for FD systems [31] [33].

4) Low-complexity algorithms are proposed for EE maximization. A channel gain based DA clustering algorithm is first performed to activate/deactivate transmit/receive chains, which highlights the characteristics of the distributed deployment, and a distributed hybrid duplexing (Dis-Hyb-Duplexing) algorithm is then performed to optimize the downlink beamformer and the uplink transmission power.

Notations: Matrices and vectors are represented by boldface capital and lower case letters, respectively. $|\cdot|$ denotes the absolute value of a complex scalar. $||\cdot||$ denotes the Euclidean vector norm. $\boldsymbol{A}^H$ and Tr($\boldsymbol{A}$) denote the Hermitian transpose and trace of matrix $\boldsymbol{A}$. Rank($\boldsymbol{A}$) denote the rank of matrix $\boldsymbol{A}$. diag ($\boldsymbol{A}$) returns a diagonal matrix with diagonal elements from matrix $\boldsymbol{A}$. $\boldsymbol{A} \succeq 0$ means $\boldsymbol{A}$ is a positive semi-definite matrix. Superscript k or u denote the downlink/uplink users' indexes.

II. System Model and Problem Formulation

A. System Model

We consider a bi-directional DA system with one central signal processing unit, which consists of a baseband module and L DAs. All DA ports are connected to the central unit through a noise-free wired front-haul for cooperative communications. The DAs can work in FD, HD (transmit or receive) or sleep mode, while the users employ single-antenna HD for low hardware complexity. In particular, there are U uplink users and K downlink users. Channel state information (CSI) is obtained by channel estimation in the training phase for both uplink and downlink, based on channel reciprocity as in bi-directional FD MIMO systems [34] [35] [36]. We assume that all the downlink/uplink users and DAs utilize orthogonal pilot sequences, which allows the DAs to distinguish the pilots of different users. In the training phase, the uplink users send pilots to the DAs and uplink CSI can be estimated at the DAs. The downlink CSI can be obtained through the channel quality indicator from downlink users [34]. Hence, the DAs

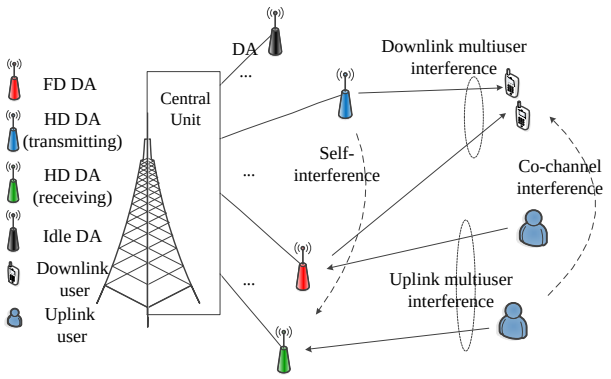

Fig. 1. Bidirectional hybrid duplexing DA system, with multiple uplink/downlink users.

can regularly update the CSI estimates of the uplink and downlink channels [34] [36]. Also, each DA can send pilots to other DAs, and thus self-interference matrix can be obtained at all DAs. For the co-channel interference channels, since the downlink users can receive the pilots from the uplink users and feed back the estimates for DAs, the DAs can update the CSI of the co-channel interference channels as well [34]. The system model is depicted by Fig. 1.

Define a receive chain activation vector $\boldsymbol{r} \in \mathbb{C}^{L\times 1}$, whose l-th element r_l equals to 1 if the l-th DA's receive chain is activated, otherwise r_l equals to 0. Define a transmit chain activation vector $\boldsymbol{t} \in \mathbb{C}^{L\times 1}$, whose l-th element t_l equals to 1 if the l-th DA's transmit chain is activated, otherwise t_l equals to 0. The configuration is reasonable in practice, since the shared-antenna deployment has been extensively researched, where only one antenna set is adopted for simultaneous transmission and reception, and the received and transmitted signals are separated with the help of a duplexer [37] [38]. Define $\boldsymbol{w_k} \in \mathbb{C}^{L\times 1}$ as beamformer at DAs for downlink user $\forall k \in K$. For each vector $\boldsymbol{w_k}$, its element w_k^l denotes the beamforming weight of DA l for user k. Denote p_u as transmission power allocation at uplink users, for $\forall u \in U$.

B. Problem Formulation

To maximize the system EE, we jointly optimize DA vectors $\boldsymbol{r}, \boldsymbol{t}$, beamformer $\boldsymbol{w_k}$, for $\forall k \in K$, and transmission power p_u, $\forall u \in U$. Define $\omega(\boldsymbol{r}, \boldsymbol{t}, \boldsymbol{w_k}, p_u)$ as the EE (in bits/Joule/Hz), which is the ratio of the system total throughput T_{total} to the incurred total power consumption P_{total}. Accordingly, the EE maximization problem is formulated as

$$\begin{aligned} P1: & \underset{\boldsymbol{r},\boldsymbol{t},\boldsymbol{w_k},p_u,k\in K,u\in U,l\in N}{\text{argmax}} \frac{T_{\text{total}}}{P_{\text{total}}}, \\ s.t.\ & (C1): 0 \le \sum_{k=1}^{K} |w_k^l|^2 \le p_{DA}, \forall l \in L, \\ & (C2): 0 \le p_u \le p_{u,max}, \forall u \in U, \\ & (C3): r_l = \{0,1\}, \forall l \in L, \\ & (C4): t_l = \{0,1\}, \forall l \in L, \end{aligned} \tag{1}$$

where $(C1)$ denotes that the allocated transmission power at each DA is non-negative and upper bounded by p_{DA}, for $\forall l \in L$. $(C2)$ denotes that the transmission power allocated at each uplink user u is non-negative and upper bounded by $p_{u,max}$, for $\forall u \in U$. $(C3)$ and $(C4)$ denote the statuses of the receive and transmit chains of DA l, for $\forall l \in L$.

III. Throughput and Power Consumption Analysis

The optimization problem involves with the throughput and power consumption in both downlink and uplink. In this section, we first analyze the downlink and uplink throughput in Subsection III-A and III-B, respectively, and then give the total power consumption in Subsection III-C.

A. Downlink Throughput

In each time slot, the received signal at downlink user k is given by

$$\begin{aligned} y_k^{DL} = \boldsymbol{h_k}(\boldsymbol{t} \circ \boldsymbol{w_k}) d_k^{DL} + \underbrace{\sum_{k'\neq k}^{K} \boldsymbol{h_k}(\boldsymbol{t} \circ \boldsymbol{w_{k'}}) d_{k'}^{DL}}_{\text{multiuser interference}} \\ + \underbrace{\sum_{u=1}^{U} \sqrt{p_u} e_{u,k} d_u^{UL}}_{\text{co-channel interference}} + z_k, \end{aligned} \tag{2}$$

where the operator $\circ$ denotes Hadamard product of two vectors or matrices. Vector $\boldsymbol{h_k} \in \mathbb{C}^{1\times L}$ is downlink channel between L DAs and downlink user k and captures the PL and small scale fading, whose l-th element $(\boldsymbol{h_k})_l$ presents the channel condition between the l-th DA and downlink user k. $d_k^{DL} \in \mathbb{C}$ is transmitted data for downlink user k. $e_{u,k} \in \mathbb{C}$ denotes the channel condition from uplink user u to downlink user k. $d_u^{UL} \in \mathbb{C}$ is the transmitted data from uplink user u. $z_k \sim CN(0, \sigma_k^2)$ is the complex Additive White Gaussian Noise (AWGN) at downlink user k. Without loss of generality, we assume $\mathbb{E}\{|d_k^{DL}|^2\} = \mathbb{E}\{|d_u^{UL}|^2\} = 1$, for $\forall k \in K$ and $\forall u \in U$. For simplicity, we introduce the auxiliary matrices $\boldsymbol{F_l} = \text{diag}\{\underbrace{0,...,0}_{l\text{-}1}, 1, \underbrace{0,...,0}_{L\text{-}l}\}$ to transform the Hadamard product $\boldsymbol{t} \circ \boldsymbol{w_k}$ into a simpler form $\boldsymbol{t} \circ \boldsymbol{w_k} = \sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_k}$. Therefore, equation (2) can be re-expressed as

$$\begin{aligned} y_k^{DL} = \boldsymbol{h_k}(\sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_k}) d_k^{DL} + \\ \sum_{k'\neq k}^{K} \boldsymbol{h_k}(\sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_{k'}}) d_{k'}^{DL} + \sum_{u=1}^{U} \sqrt{p_u} e_{u,k} d_u^{UL} + z_k. \end{aligned} \tag{3}$$

According to (3), the signal-to-interference-and-noise ratio (SINR) of user k is calculated as

$$\Gamma_k^{DL} = \frac{\left| \boldsymbol{h_k}(\sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_k}) \right|^2}{\sum_{k'\neq k}^{K} \left| \boldsymbol{h_k}(\sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_{k'}}) \right|^2 + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2}. \tag{4}$$

Thus, the total downlink throughput is calculated as

$$T_{DL} = \sum_{k=1}^{K} \log_2(1 + \Gamma_k^{DL}). \tag{5}$$

Substituting (4) in (5) and transforming (5) into the structure of $\log(1+\frac{A}{B}) = \log(A+B) - \log B$, we have

$$\begin{aligned} T_{DL} = & \\ & \sum_{k=1}^{K} \log_2\Big(\sum_{k'=1}^{K} |\boldsymbol{h_k}(\sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_{k'}})|^2 + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \Big) - \\ & \sum_{k=1}^{K} \log_2\Big(\sum_{k'\neq k}^{K} |\boldsymbol{h_k}(\sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_{k'}})|^2 + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \Big). \end{aligned} \tag{6}$$

Defining matrices $\boldsymbol{H_k} = \boldsymbol{h_k^H h_k}$ and $\boldsymbol{W_k} = \boldsymbol{w_k w_k^H}$ and substituting them into (6), the total downlink throughput is finally given by

$$\begin{aligned} T_{DL} = & \sum_{k=1}^{K} \log_2\Big(\sum_{k'=1}^{K} \mathrm{Tr}(\sum_{m=1}^{L}\sum_{n=1}^{L} t_m t_n \boldsymbol{F_m W_{k'} F_n^H H_k}) \\ & + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \Big) \\ & - \sum_{k=1}^{K} \log_2\Big(\sum_{k'\neq k}^{K} \mathrm{Tr}(\sum_{m=1}^{L}\sum_{n=1}^{L} t_m t_n \boldsymbol{F_m W_{k'} F_n^H H_k}) \\ & + \sum_{u=1}^{U} p_u |e_{u,k}|^2 + \sigma_k^2 \Big). \end{aligned} \tag{7}$$

B. Uplink Throughput

In uplink, define self-interference channel matrix $\boldsymbol{H_{SI}} \in \mathbb{C}^{L\times L}$, whose element $(\boldsymbol{H_{SI}})_{ij}$ represents the channel condition from the j-th DA to the i-th DA. Specifically, the diagonal entry $(\boldsymbol{H_{SI}})_{ii}$, for $\forall i \in L$, represents the i-th DA's self-interference channel. The self-interference signal is the downlink transmitted signal from the DAs, and thus can be calculated as $\sum_{k=1}^{K}\sum_{l=1}^{L} t_l \boldsymbol{F_l w_k} d_k^{DL}$. Define an SIC amount α as the ratio of post-SIC self-interference power over the pre-SIC self-interference power. The residual self-interference after SIC is thus given by $\frac{1}{\sqrt{\alpha}} \boldsymbol{H_{SI}} \sum_{k=1}^{K}\sum_{l=1}^{L} t_l \boldsymbol{F_l w_k} d_k^{DL}$, and the uplink received signal at the DAs is given by

$$\boldsymbol{y}^{UL} = \sum_{u=1}^{U} \sqrt{p_u} \boldsymbol{g_u} d_u^{UL} + \underbrace{\frac{1}{\sqrt{\alpha}} \boldsymbol{H_{SI}} \sum_{k=1}^{K}\sum_{l=1}^{L} t_l \boldsymbol{F_l w_k} d_k^{DL}}_{\text{residual self-interference}} + \boldsymbol{z}, \tag{8}$$

where vector $\boldsymbol{g_u} \in \mathbb{C}^{L\times 1}$ is the uplink channel from user u to L DAs, whose element $(\boldsymbol{g_u})_l$ presents the channel condition between the l-th DA and uplink user u. Vector $\boldsymbol{z} \in \mathbb{C}^{L\times 1}, \sim CN(0, \sigma^2 \boldsymbol{I_L})$ is the AWGN receive noise. We assume that the central unit employs a linear maximum ratio combining (MRC) receiver [31] [1], $\boldsymbol{v_u} \in \mathbb{C}^{L\times 1}$, for decoding of the received u-th uplink user information, which is given by $\boldsymbol{v_u} = \sum_{i=1}^{L} r_i \boldsymbol{F_i g_u}$. It is obvious that $r_i = 0$ means the receive chain of the i-th DA is deactivated. As a result, the equivalent SINR for uplink user u is given by

$$\Gamma_u^{UL} = \frac{p_u |\boldsymbol{v_u^H g_u}|^2}{\sum_{u'\neq u}^{U} p_{u'} |\boldsymbol{v_u^H g_{u'}}|^2 + \frac{\left|\boldsymbol{v_u^H H_{SI}} \sum_{k=1}^{K}\sum_{l=1}^{L} t_l \boldsymbol{F_l w_k}\right|^2}{\alpha} + ||\boldsymbol{v_u}||^2 \sigma^2}. \tag{9}$$

Therefore, the total uplink throughput is given by

$$T_{UL} = \sum_{u=1}^{U} \log_2(1 + \Gamma_u^{UL}). \tag{10}$$

Substituting (9) into (10) leads

$$\begin{aligned} T_{UL} = & \sum_{u=1}^{U} \log_2\Big(\sum_{u'=1}^{U} p_{u'} |\boldsymbol{v_u^H g_{u'}}|^2 + \\ & \frac{|\boldsymbol{v_u^H H_{SI}} \sum_{k=1}^{K}\sum_{l=1}^{L} t_l \boldsymbol{F_l w_k}|^2}{\alpha} + ||\boldsymbol{v_u}||^2 \sigma^2 \Big) \\ & - \sum_{u=1}^{U} \log_2\Big(\sum_{u'\neq u}^{U} p_{u'} |\boldsymbol{v_u^H g_{u'}}|^2 + \\ & \frac{|\boldsymbol{v_u^H H_{SI}} \sum_{k=1}^{K}\sum_{l=1}^{L} t_l \boldsymbol{F_l w_k}|^2}{\alpha} + ||\boldsymbol{v_u}||^2 \sigma^2 \Big). \end{aligned} \tag{11}$$

Substituting $\boldsymbol{G_u} = \boldsymbol{g_u g_u^H}$ and $\boldsymbol{v_u} = \sum_{i=1}^{L} r_i \boldsymbol{F_i g_u}$ into (11), the total uplink throughput is given by (12). Removing the quadratic terms in (12), (12) is finally derived into (13).

C. Power Consumption

As researched by [39] [40] [41], the power consumption in DA systems mainly includes PA power consumption, circuit power consumption and system fixed power consumption [19] [23] [42]. Especially with FD technique, additional power may be incurred for SIC operation.

a) It was pointed out in the FP7 EARTH project [43] that for a small-scale communication node (such DA, femto or pico node), its PA power accounts for 40%-47% of the total power [43]. The PA power is closely related to the beamforming weights at DAs, uplink transmission power at users and drain efficiency (DE) of PA, which is given by $\frac{1}{\eta}\big(\mathrm{Tr}(\sum_{k=1}^{K} \boldsymbol{W_k}) + \sum_{u=1}^{U} p_u\big)$ and η is the DE of PAs. Without loss of generality, we assume that all PAs (of DAs and users) work in linear region and have the same DE performance [23]. Besides, per-DA and per-uplink user power constraints are applied. This is because the DAs and users have individual transmit chain and cannot share power with each other.

[1]It is known that MRC provides a performance close to minimum mean square error (MMSE) beamforming given a large number of antennas [44]. Besides, since the coupling between $\boldsymbol{r}, \boldsymbol{t}, \boldsymbol{W_k}$ and the residual self-interference at DAs, it is difficult to derive a tractable solution by applying MMSE or zero-forcing receive beamforming.

$$
\begin{aligned}
T_{UL} = &\sum_{u=1}^{U} \log_2\Bigg(\sum_{u'=1}^{U} p_{u'} \left| (\sum_{i=1}^{L} r_i \boldsymbol{F_i} \boldsymbol{g_u})^H \boldsymbol{g_{u'}} \right|^2 + \frac{1}{\alpha} \left| (\sum_{i=1}^{L} r_i \boldsymbol{F_i} \boldsymbol{g_u})^H \boldsymbol{H_{SI}} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_k} \right|^2 + ||(\sum_{i=1}^{L} r_i \boldsymbol{F_i} \boldsymbol{g_u})^H||^2 \sigma^2 \Bigg) \\
&- \sum_{u=1}^{U} \log_2\Bigg(\sum_{u' \neq u}^{U} p_{u'} \left| (\sum_{i=1}^{L} r_i \boldsymbol{F_i} \boldsymbol{g_u})^H \boldsymbol{g_{u'}} \right|^2 + \frac{1}{\alpha} \left| (\sum_{i=1}^{L} r_i \boldsymbol{F_i} \boldsymbol{g_u})^H \boldsymbol{H_{SI}} \sum_{k=1}^{K} \sum_{l=1}^{L} t_l \boldsymbol{F_l} \boldsymbol{w_k} \right|^2 + ||(\sum_{i=1}^{L} r_i \boldsymbol{F_i} \boldsymbol{g_u})^H||^2 \sigma^2 \Bigg).
\end{aligned} \tag{12}
$$

$$
\begin{aligned}
T_{UL} = &\sum_{u=1}^{U} \log_2\Bigg(\sum_{u'=1}^{U} p_{u'} \mathrm{Tr}(\boldsymbol{G_{u'}} \sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j \boldsymbol{F_i} \boldsymbol{G_u} \boldsymbol{F_j}^H) + \sigma^2 \mathrm{Tr}(\sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j \mathbf{F}_i \mathbf{G}_u \mathbf{F}_j^H) + \\
&\frac{1}{\alpha} \mathrm{Tr}\Big(\mathrm{diag}\big(\boldsymbol{H_{SI}} (\sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m} \boldsymbol{W_k} \boldsymbol{F_n}^H) \boldsymbol{H_{SI}}^H \big) (\sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j \boldsymbol{F_i} \boldsymbol{G_u} \boldsymbol{F_j}^H) \Big) \Bigg) \\
&- \sum_{u=1}^{U} \log_2\Bigg(\sum_{u' \neq u}^{U} p_{u'} \mathrm{Tr}(\boldsymbol{G_{u'}} \sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j \boldsymbol{F_i} \boldsymbol{G_u} \boldsymbol{F_j}^H) + \sigma^2 \mathrm{Tr}(\sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j \boldsymbol{F_i} \boldsymbol{G_u} \boldsymbol{F_j}^H) + \\
&\frac{1}{\alpha} \mathrm{Tr}\Big(\mathrm{diag}\big(\boldsymbol{H_{SI}} (\sum_{k=1}^{K} \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m} \boldsymbol{W_k} \boldsymbol{F_n}^H) \boldsymbol{H_{SI}}^H \big) (\sum_{i=1}^{L} \sum_{j=1}^{L} r_i r_j \boldsymbol{F_i} \boldsymbol{G_u} \boldsymbol{F_j}^H) \Big) \Bigg)
\end{aligned} \tag{13}
$$

b) Circuit power consumption is proportional to the number of active transmit and receive chains [39]. Transmit chain and receive chains contain multiple components, such as digital/analog (D/A) converter, analog/digital (A/D) converter, electrical/optical (E/O) converter, optical/electrical (O/E) converter, up-converter, down-converter, filter, synthesizer, mixers, *etc* [23]. Therefore, the power consumption of transmit chain and receive chains can be formulated by $p_{r,t}$ and $p_{c,t}$, respectively. When both transmit and receive chains are deactivated, the DA is in sleep mode and the power consumption is greatly reduced to p_{idle} based on the state-of-the-art switch-off technique [31]. Therefore, the total circuit power consumption is calculated as $\sum_{l=1}^{L} \big((1-r_l)(1-t_l)p_{idle}\big) + \sum_{l=1}^{L}(r_l p_{c,r}) + \sum_{l=1}^{L}(t_l p_{c,t})$.

c) System fixed power consumption p_{fix} is the power consumed by power supply, active cooling system at central unit and/or DAs, *etc*., which is independent of the state of transmit/receive chain [31].

d) For the DAs working in FD mode, *i.e*., $r_l = 1$ and $t_l = 1$, for $\forall l \in L$, additional power is required for SIC, which is related to the specific SIC circuit design [2], and can be considered as a constant value [45] [46]. It is because passive suppression benefits from PL and antenna isolation, which actually do not consume additional power. While the power of analog and digital cancellation is non-negligible compared to passive suppression. The power consumed by the involved components, such as filter, attenuator, splitter and adder can be formulated as constant [18]. Define p_{can} as the power required by SIC at each DA, and the overall power consumed by SIC is given as $\sum_{l=1}^{L}(r_l t_l p_{can})$ [3].

Finally, the total power consumption is given by

$$
\begin{aligned}
P_{\text{total}} = &\underbrace{\frac{1}{\eta}\Big(\mathrm{Tr}(\sum_{k=1}^{K} \boldsymbol{W_k}) + \sum_{u=1}^{U} p_u \Big)}_{\text{PA power}} + \underbrace{\sum_{l=1}^{L}(t_l r_l p_{can})}_{\text{SIC power}} + p_{fix} + \\
&\underbrace{\sum_{l=1}^{L} \Big((1-r_l)(1-t_l) p_{idle} \Big) + \sum_{l=1}^{L}(r_l p_{c,r}) + \sum_{l=1}^{L}(t_l p_{c,t})}_{\text{circuit power}}.
\end{aligned} \tag{14}
$$

IV. Energy Efficiency (EE) Performance Optimization

According to the throughput and power consumption analyzed in Section III, the problem $P1$ is re-formulated into

$$
\begin{aligned}
&P2: \\
&\underset{\boldsymbol{r},\boldsymbol{t},\boldsymbol{W_k},p_u,k\in K,u\in U,l\in L}{\mathrm{argmax}} \frac{T_{DL}(\boldsymbol{r},\boldsymbol{t},\boldsymbol{W_k},p_u) + T_{UL}(\boldsymbol{r},\boldsymbol{t},\boldsymbol{W_k},p_u)}{P_{\text{total}}(\boldsymbol{r},\boldsymbol{t},\boldsymbol{W_k},p_u)}, \\
&s.t.\ (\widetilde{C1}): \sum_{k=1}^{K} \mathrm{Tr}(\boldsymbol{W_k} \boldsymbol{F_l}) \leq p_{DA}, \forall l \in L, \\
&\quad (C2), (C3), (C4), \\
&\quad (C5): \boldsymbol{W_k} \succeq 0, \forall k \in K, \\
&\quad (C6): \mathrm{Rank}(\boldsymbol{W_k}) = 1, \forall k \in K,
\end{aligned} \tag{15}
$$

[2]Generally speaking, complex SIC scheme needs more involved components and consumes higher power. For example, the analog cancellation design in [1] needs D/A converter, transmit radio unit and adder to mitigate self-interference. Another kind of design in [17] uses tunable attenuator and delay unit to route the estimated signal to the receiver for SIC.

[3]Herein, analog or digital of SIC is only required within each FD DA itself, which is because the self-interference among different DAs can be well controlled in the propagation domain benefiting from the high PL. Also, SIC operation across different DAs imposes high complexity for hardware design. More details can be referred in [1] for co-located FD MIMO systems.

where constraint $(\widetilde{C1})$ denotes that the transmission power at each DA is upper bounded by p_{DA}, for $\forall l \in L$. $(C5)$ and $(C6)$ are imposed to guarantee that $\boldsymbol{W_k} = \boldsymbol{w_k}\boldsymbol{w_k}^H$ holds after optimization, for $\forall k \in K$.

The problem in $P2$ is generally very difficult to solve. It involves a) vectors $\boldsymbol{r}, \boldsymbol{t}$ with binary elements in the objective function and the combinational constraints in $(C3), (C4)$. b) product of binary variables t_m, t_n, r_i, r_j with continuous variables $\boldsymbol{W_k}, p_u$, for $\forall k \in K, u \in U$ and $\forall m, n, i, j \in L$. c) rank constraint in $(C6)$. Besides, with L DAs, there are 4^L possibilities of DA configurations. It is prohibitive to find the global optimum in terms of computational complexity and thus a low-complexity suboptimal design is desirable. In the Subsection IV-A, we first propose a channel-gain-based DA clustering algorithm to perform DAs activation/deactivation, which effectively removes the binary variables in the objective and in the constraints $(C3)$ $(C4)$, as well as the product between binary variables with continuous variables. After this, a novel distributed hybrid duplexing (Dis-Hyb-Duplexing) algorithm is proposed to optimize beamformer $\boldsymbol{W_k}$, for $\forall k \in K$ and transmission power p_u, for $\forall u \in U$, in Subsection IV-B. At last, total complexity is given in IV-C.

A. Activation/Deactivation of Transmit/Receive Chains at DAs

Intuitively, if the channel gain between the l-th DA and uplink users is strong, while the channel gain between the l-th DA and downlink users is poor, the l-th DA should work in HD receive mode. Since if the DA works in FD mode, its downlink throughput contribution is marginal for all downlink users, while more power consumption is required. More importantly, the downlink transmission corrupts its uplink reception due to the introduced self-interference. Conversely, the DA should work in HD transmit mode, if the channel gain between the DA and uplink users is poor while the channel gain between the DA and downlink users is strong. Also, one DA can work in FD mode if it has good channel condition in both uplink and downlink, which contributes reasonable throughput bi-directionally with affordable power consumption. At last, the DA can be turned off to save power consumption if it has poor channel condition from all users.

To implement the DA clustering algorithm, a threshold, *i.e.,* ψ, is needed for judgment. The threshold can be adjusted according to different quality of service (QoS) requirements and densities of DAs. For example, one can increase ψ if power consumption requirement is stringent. It means less DAs will be activated and therefore power consumption is decreased. If users' SE requirement is stringent, on the other hand, one can decrease the value of ψ, which means more DAs will be activated. Also, for a dense DA deployment, the value of ψ could be higher than a sparse deployment. The whole DA clustering algorithm is presented in Algorithm 1, then downlink beamformer and uplink transmission power allocation are ready to perform.

Algorithm 1 DA clustering algorithm

Input: $\boldsymbol{h_k}, \boldsymbol{g_u}$ for $\forall k \in k, u \in U$, and the threshold ψ.
Output: Transmit/receive chains vector $\boldsymbol{r}^*$, $\boldsymbol{t}^*$.

```
for l = 1, ..., L do
    for u = 1, ..., U do
        Calculate the channel condition between the l-th DA and uplink user u.
        if (g_u)_l ≥ ψ then
            r_l = 1, break.
        else
            r_l = 0.
        end if
    end for
    for k = 1, ..., K do
        Calculate the channel condition between the l-th DA and downlink user k.
        if (h_k)_l ≥ ψ then
            t_l = 1, break.
        else
            t_l = 0.
        end if
    end for
end for
return r*, t*.
```

B. Design of Downlink Beamformer at DAs and Uplink Transmission Power at Users

For simplicity, we define a super matrix $\boldsymbol{W} = \{\boldsymbol{W_1}, \boldsymbol{W_2}, ..., \boldsymbol{W_K}\}$, including all beamformer variables $\boldsymbol{W_k}, \forall k \in K$. Define a vector $\boldsymbol{p} = \{p_1, p_2, ..., p_U\}$, including all uplink transmission power variables $p_u, \forall u \in U$. The feasible domain confined by the constraints is expressed as $\{\boldsymbol{\Theta}\}$. Then the optimization problem goes into

$$P3: \underset{\boldsymbol{W},\boldsymbol{p}\in\{\boldsymbol{\Theta}\}}{\operatorname{argmax}} \frac{T_{DL}(\boldsymbol{W},\boldsymbol{p}) + T_{UL}(\boldsymbol{W},\boldsymbol{p})}{P_{\text{total}}(\boldsymbol{W},\boldsymbol{p})}, \quad s.t.\ (\widetilde{C1}), (C2), (C5) \text{ and } (C6). \tag{16}$$

For the total throughput $T_{DL}(\boldsymbol{W},\boldsymbol{p}) + T_{UL}(\boldsymbol{W},\boldsymbol{p})$, we collect the positive parts into $f_1(\boldsymbol{W},\boldsymbol{p})$, and the negative parts into $f_2(\boldsymbol{W},\boldsymbol{p})$. Then the total throughput can be expressed as $f_1(\boldsymbol{W},\boldsymbol{p}) - f_2(\boldsymbol{W},\boldsymbol{p})$, where $f_1(\boldsymbol{W},\boldsymbol{p})$ and $f_2(\boldsymbol{W},\boldsymbol{p})$ are given by (17) and (18), respectively.

Obviously, $f_1(\boldsymbol{W},\boldsymbol{p})$ and $f_2(\boldsymbol{W},\boldsymbol{p})$ are both jointly-concave with respect to the variables $\boldsymbol{W},\boldsymbol{p}$ in the considered domain, and the Frank-Wolfe (FW) method can be adopted to handle the difference of two concave functions. The FW method approximates the minus part $f_2(\boldsymbol{W},\boldsymbol{p})$ by its first order Taylor series $f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})$, and updates the approximation $f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})$ iteratively along the direction that approaches the original function $f_2(\boldsymbol{W},\boldsymbol{p})$ [11]. Suppose the value of $\boldsymbol{W},\boldsymbol{p}$ is denoted by $\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}}$ and $f_2(\boldsymbol{W},\boldsymbol{p})$ is approximated by $f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})$ at the n-th iteration. Since the total throughput $f_1(\boldsymbol{W},\boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})$ is the lower bound of the original one $f_1(\boldsymbol{W},\boldsymbol{p}) - f_2(\boldsymbol{W},\boldsymbol{p})$, the variables $\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}}$ are iteratively updated and the lower bound increases after every iteration. Because of the power consumption is upper bounded by the constraints, the iterative procedure is guaranteed to converge. Since $f_2(\boldsymbol{W},\boldsymbol{p})$ is concave and differentiable on the considered domain, one can easily find its first-order approximation, as

$$f_1(\boldsymbol{W},\boldsymbol{p})=\sum_{k=1}^{K}\log_2\Bigg(\sum_{k'=1}^{K}\mathrm{Tr}(\sum_{m=1}^{L}\sum_{n=1}^{L}t_mt_n\boldsymbol{F_m}\boldsymbol{W_{k'}}\boldsymbol{F_n}^H\boldsymbol{H_k})+\sum_{u=1}^{U}p_u|e_{u,k}|^2+\sigma_k^2\Bigg)+ \sum_{u=1}^{U}\log_2\Bigg(\sum_{u'=1}^{U}\Big(p_{u'}\mathrm{Tr}(\boldsymbol{G_{u'}}\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)\Big)+\sigma^2\mathrm{Tr}(\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)+ \frac{1}{\alpha}\mathrm{Tr}\Big(\mathrm{diag}\big(\boldsymbol{H_{SI}}(\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}t_mt_n\boldsymbol{F_m}\boldsymbol{W_k}\boldsymbol{F_n}^H)\boldsymbol{H_{SI}}^H\big)(\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)\Big)\Bigg). \tag{17}$$

$$f_2(\boldsymbol{W},\boldsymbol{p})=\sum_{k=1}^{K}\log_2\Bigg(\sum_{k'\neq k}^{K}\mathrm{Tr}(\sum_{m=1}^{L}\sum_{n=1}^{L}t_mt_n\boldsymbol{F_m}\boldsymbol{W_{k'}}\boldsymbol{F_n}^H\boldsymbol{H_k})+\sum_{u=1}^{U}p_u|e_{u,k}|^2+\sigma_k^2\Bigg)+ \sum_{u=1}^{U}\log_2\Bigg(\sum_{u'\neq u}^{U}\Big(p_{u'}\mathrm{Tr}(\boldsymbol{G_{u'}}\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)\Big)+\sigma^2\mathrm{Tr}(\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)+ \frac{1}{\alpha}\mathrm{Tr}\Big(\mathrm{diag}\big(\boldsymbol{H_{SI}}(\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}t_mt_n\boldsymbol{F_m}\boldsymbol{W_k}\boldsymbol{F_n}^H)\boldsymbol{H_{SI}}^H\big)(\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)\Big)\Bigg). \tag{18}$$

$$\begin{aligned}&f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})=f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})+\\ &\sum_{k=1}^{K}\Bigg((a_k^{(n)})^{-1}\sum_{k'\neq k}^{K}\mathrm{Tr}(\sum_{m=1}^{L}\sum_{n=1}^{L}t_mt_n\boldsymbol{F_m}(\boldsymbol{W_{k'}}-\boldsymbol{W_{k'}^{(n)}})\boldsymbol{F_n}^H\boldsymbol{H_k})\Bigg)+\sum_{k=1}^{K}\Bigg((a_k^{(n)})^{-1}\sum_{u=1}^{U}(p_u-p_u^{(n)})|e_{u,k}|^2\Bigg)+\\ &\sum_{u=1}^{U}\Bigg((b_u^{(n)})^{-1}\frac{1}{\alpha}\mathrm{Tr}\Big(\mathrm{diag}\big(\boldsymbol{H_{SI}}(\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}t_mt_n\boldsymbol{F_m}(\boldsymbol{W_k}-\boldsymbol{W_k^{(n)}})\boldsymbol{F_n}^H)\boldsymbol{H_{SI}}^H\big)(\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j}^H)\Big)\Bigg)+\\ &\sum_{u=1}^{U}\Bigg((b_u^{(n)})^{-1}\sum_{u'\neq u}^{U}(p_{u'}-p_{u'}^{(n)})\mathrm{Tr}\big(\boldsymbol{G_{u'}}\sum_{i=1}^{L}\sum_{j=1}^{L}r_ir_j\boldsymbol{F_i}\boldsymbol{G_u}\boldsymbol{F_j^H}\big)\Bigg),\end{aligned} \tag{19}$$

$$(a_k^{(n)})=\sum_{k'\neq k}^{K}\Bigg(\mathrm{Tr}(\sum_{i=1}^{L}\sum_{j=1}^{L}t_mt_n\boldsymbol{F_m}\boldsymbol{W_{k'}^{(n)}}\boldsymbol{F_n}^H\boldsymbol{H_k})\Bigg)+\sum_{u=1}^{U}p_u^{(n)}|e_{u,k}|^2+\sigma_k^2, \tag{20}$$

shown by (19) on the next page. The related variables $(a_k^{(n)})$ and $(b_u^{(n)})$ in (19) are detailed by (20) and (21), respectively.

Now, the overall throughput is re-expressed as $f_1(\boldsymbol{W},\boldsymbol{p})-f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})$, which is jointly-concave with respect to the variables. It is because $f_1(\boldsymbol{W},\boldsymbol{p})$ is jointly-concave, and $f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})$ is affine with respect to all variables $\boldsymbol{W},\boldsymbol{p}\in\{\boldsymbol{\Theta}\}$. The optimization problem can be expressed as

$$\begin{aligned}&P4:\ \underset{\boldsymbol{W},\boldsymbol{p}\in\{\boldsymbol{\Theta}\}}{\mathrm{argmax}}\ \frac{f_1(\boldsymbol{W},\boldsymbol{p})-f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})}{P_{\mathrm{total}}(\boldsymbol{W},\boldsymbol{p})},\\ &s.t.\ (\widetilde{C1}),(C2),(C5)\text{ and }(C6).\end{aligned} \tag{22}$$

Since the problem $P4$ is the ratio between a concave function and an affine function in the consider domain, we introduce Theorem 1 to solve the problem.

Theorem 1: The reformulated EE, $\dfrac{f_1(\boldsymbol{W},\boldsymbol{p})-f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})}{P_{\mathrm{total}}(\boldsymbol{W},\boldsymbol{p})}$, is jointly quasi-concave with respect to variables $\boldsymbol{W}$ and $\boldsymbol{p}$ in the feasible domain.

Proof: See APPENDIX A.

Theorem 1 confirms that there is a global optimal EE in the feasible domain. However, the objective function of (22) is in a fractional structure, which is still difficult to handle. Therefore, we further introduce Proposition 1 to transfer (22) into an equivalent but easier structure.

Proposition 1: Assume that β^* is the maximum value of $\underset{\boldsymbol{W},\boldsymbol{p}\in\{\boldsymbol{\Theta}\}}{\mathrm{argmax}}\ \frac{f_1(\boldsymbol{W},\boldsymbol{p})-f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})}{P_{\mathrm{total}}(\boldsymbol{W},\boldsymbol{p})}$ in (22). Maximizing the quasi-concave problem in (22) is equivalent to finding the root of a subtract programming problem $f_1(\boldsymbol{W},\boldsymbol{p})-f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})-\beta^*P_{\mathrm{total}}(\boldsymbol{W},\boldsymbol{p})$ [47] [48].

Proof: See APPENDIX B.

According Proposition 1, the problem in (22) can be associated with an equivalent subtract programming problem. Thus solving the problem $P4$ turns into solving $f_1(\boldsymbol{W},\boldsymbol{p})-$

$$(b_u^{(n)}) = \sigma^2 \mathrm{Tr}(\sum_{i=1}^{L}\sum_{j=1}^{L} r_i r_j \boldsymbol{F_i G_u F_j}^H) + \sum_{u' \neq u}^{U} \left(p_{u'}^{(n)} \mathrm{Tr}(\sum_{i=1}^{L}\sum_{j=1}^{L} r_i r_j \boldsymbol{F_i G_u F_j}^H) \right) + \frac{1}{\alpha} \mathrm{Tr}\left(\mathrm{diag}\left(\boldsymbol{H_{SI}} (\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L} t_m t_n \boldsymbol{F_m W_k}^{(n)} \boldsymbol{F_n}^H) \boldsymbol{H_{SI}}^H \right) (\sum_{i=1}^{L}\sum_{j=1}^{L} r_i r_j \boldsymbol{F_i G_u F_j}^H) \right). \tag{21}$$

$f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p})$ with a given β, as shown by the problem $P5$ in (23).

$$\begin{aligned} P5: \ & \underset{\boldsymbol{W}, \boldsymbol{p} \in \{\boldsymbol{\Theta}\}}{\text{argmax}} \ f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}), \\ & s.t. \ (\widetilde{C1}), (C2), (C5), (C6). \end{aligned} \tag{23}$$

Now we introduce Proposition 2 to handle the rank one constraint in $(C6)$.

Proposition 2: For the rank constraint in $(C6)$, rank relaxation can be applied [34] [47] [49]. In particular, with the condition that the channel parameters are statistically independent, the solution to the original problem is of rank-one.

Proof: See APPENDIX C.

In DA systems, the channel is independent [23], which is different from the co-located antennas systems. According to Proposition 2, we drop the rank-one constraint in $(C6)$ benefiting from the DA deployment. Now the problem $P5$ after rank-one relaxation is a standard concave maximization problem with semi-definite programming (SDP). Therefore, we can adopt the CVX solver to solve the problem [4]. Finally, a Dis-Hyb-Duplexing algorithm is proposed to concurrently optimize $\boldsymbol{W}$ and $\boldsymbol{p}$, as summarized by Algorithm 2. The convergence behavior of the distributed hybrid duplexing algorithm is summarized in Proposition 3.

Proposition 3: The proposed Dis-Hyb-Duplexing algorithm solves the SDP problem $(P5)$ in the inner layer and updates β in the outer layer. Finally, optimal beamformer $\boldsymbol{W}^*$ and uplink transmission power $\boldsymbol{p}^*$ are readily obtained.

Proof: See APPENDIX D.

Algorithm 2 Dis-Hyb-Duplexing algorithm

Input: Antenna configuration vectors $\boldsymbol{r}, \boldsymbol{t}$, left/right bounds β_l and β_r, channel condition, *i.e.,* $\boldsymbol{h_k}, \boldsymbol{g_u}, e_{u,k}, \boldsymbol{h_{SI}}$, for $\forall k \in k, u \in U$, and power consumption parameters, *i.e.,* $\eta, p_{idle}, p_{c,t}, p_{c,r}, p_{can}, p_{fix}, p_{DA}, p_{u,max}$.

Output: Optimal beamforming weight $\boldsymbol{W_k^*}$, for $\forall k \in k$, and optimal uplink transmission power p_u^*, $\forall u \in U$.

Set accuracy factor $\epsilon > 0$, and suppose $\mathcal{F}(\beta)$ is the optimal value of $f_1(\boldsymbol{W}, \boldsymbol{p}) + f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p})$. Initialize left bound β_l and right bound β_r that ensure the $\mathcal{F}(\beta_l) \cdot \mathcal{F}(\beta_r) < 0$.
while $\beta_r - \beta_l > \epsilon$ **do**
  $\beta = \frac{\beta_r + \beta_l}{2}$.
  Solve the problem $P5$ with the FW method until convergence.
  if $\mathcal{F}(\beta_l) \cdot \mathcal{F}(\beta) < 0$ **then**
    $\beta_r = \beta$.
  else
    $\beta_l = \beta$.
  end if
end while

C. Complexity Analysis

In this subsection, we first analyze the complexity of the two algorithms and then give the overall complexity. In the DA clustering algorithm, we need at most $L(K+U)$ comparisons to find $\boldsymbol{r}^*$ and $\boldsymbol{t}^*$. After this, the Dis-Hyb-Duplexing algorithm is performed to allocate beamformer at the DAs and transmission power at the uplink users. Since we transform the fractional-structure $\frac{f_1(\boldsymbol{W},\boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W},\boldsymbol{p})}{P_{\text{total}}(\boldsymbol{W},\boldsymbol{p})}$ into a more tractable difference-structure $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p})$, bisection search is used in the outer layer for finding optimal β^*. The bisection method [18] is a simple and robust root-finding method, which repeatedly bisects an interval and then selects a subinterval in which a root must lie for further processing. The bisection method is guaranteed to converge to a root of a function if the function is continuous between the left and right bounds, and the values of the functions have opposite signs at the two bounds.

Assume that $\beta_1 = \frac{\beta_r + \beta_l}{2}$ is the midpoint of the initial interval, and β_n is the midpoint of the interval in the n-th step. Then the difference between β_n and β^* is bounded by $|\beta_n - \beta^*| \leq \frac{\beta_r - \beta_l}{2^n}$. With a tolerance factor ϵ, the required number of iterations is given by $n \leq log_2(\frac{\beta_r - \beta_l}{\epsilon})$. In the proposed algorithm, the left bound β_l can be set to 0, with which the value of equation $\mathcal{F}(\beta_l)$ is definitely positive. Also, a sufficiently large value of β_r can be chosen as the right bound to make the value of $\mathcal{F}(\beta_r)$ negative. Therefore, function $\mathcal{F}(\cdot)$ has opposite signs at the two bounds and thus the bisection approach readily leads to convergence. In the inner layer, CVX is called to solve the SDP optimization, whose complexity is indexed as ξ [5]. Thus the complexity of the Dis-Hyb-Duplexing algorithm is $\mathcal{O}(\log_2(\frac{\beta_r - \beta_l}{\epsilon}) \cdot \xi)$. As a result, the total complexity is given by $\mathcal{O}\big(L(K+U) + \log_2(\frac{\beta_r - \beta_l}{\epsilon}) \cdot \xi\big)$. As can be seen, the associated bisection search in the outer

[4]The CVX contains multiple solvers for SDP optimization, *e.g.,* SEDuMi, SDPT3 and MOSEK. By claiming variables, objective function (maximization of concave function or minimization of convex function) and constraints, the SDP optimization can be readily solved.

[5]Since the FW method is adopted for the inner layer iteration, the convergence rate is $\mathcal{O}(\frac{1}{n})$ [50]. In the n-th update in the inner layer, the CVX SDPT3 solver implements an interior-point method to solve the SDP problem, which belongs to the class of path-following method and leads a fast convergence.

TABLE I. Simulation Setup

Central carrier frequency	2 GHz
Bandwidth	1 MHz
AWGN power spectral density	-174 dBm/Hz
Cell model	100 m × 100 m
DE of PA η	25%
p_{idle}, $p_{c,r}$ and $p_{c,t}$	10 mW, 100 mW and 100 mW
p_{can} and p_{fix}	50 mW and 500 mW
$p_{u,max}$	40 mW
Antenna Gain	0 dBi

layer has the largest impact on the complexity of the whole algorithm, more insight on the number of iterations of the associated bisection search will be demonstrated in Section. V.

V. Simulation Results

We use numerical results to verify our analysis, with 4000 times of Monte Carlo simulations run to obtain an average and common parameters given in TABLE I. The SIC amount is set to $\alpha = 100$ dB except for Fig. 4, where α varies from 60 to 100 dB. The DA clustering threshold is set to $\psi = 1 \times 10^{-9}$ except for Fig. 6, where ψ varies from 0.5×10^{-9} to 5×10^{-9}. The power constraint is set to $p_{DA} = 100$ mW except for Fig. 5, where p_{DA} varies from 60 mW to 140 mW. The PL model of $\mathrm{PL}(d) = 145.4 + 37.5 \log_{10}(d/1000)$ [51] is adopted, which is featured in 3GPP LTE standards at 2 GHz. The small-scale fading is modeled as Rician fading, including the self-interference channel $\boldsymbol{H_{SI}}$, with Rician factor 5 dB [52]. The DAs are uniformly deployed across an area of 100 m × 100 m, with an example of 25 DAs demonstrated in Fig. 2. 25 DAs are used for all figures except Figs. 6 and 7, where the number of DAs varies from 16 to 36. 2 uplink users with locations (92 m, 10 m) and (85 m, 86 m), respectively, and 2 downlink users with locations (5 m, 76 m) and (78 m, 16 m), respectively, are used for Figs. 3-7. While in Figs. 8 and 9, the total number of users is increased from 4 to up to 16, and their locations are independent and identically distributed (i.i.d.) random variables following uniform distribution.

The distributed sole-FD system [31] and the co-located FD MIMO system [47] are used as benchmarks. In the former, all antennas are distributed but are only capable of working in FD or sleep modes. In the latter, all antennas are centralized at the BS. Since all antennas have identical PL to users due to centralized antenna deployment, the activation/deactivation of transmit/receive chains is disabled. Hence, all antennas work in FD mode and all elements in the self-interference channel matrix $\boldsymbol{H_{SI}}$ become positive.

Fig. 3 shows the convergence behavior of the proposed Dis-Hyb-Duplexing algorithm on finding the optimal value of β^*. With left bound $\beta_l = 0$, right bound $\beta_r = 100$ and tolerance $\epsilon = 0.1$, at most 10-12 iterations are required to achieve convergence, which matches the analysis of $n \leq log_2(\frac{\beta_r - \beta_l}{\epsilon})$ iterations in Subsection IV-C and confirms the low complexity of the proposed algorithm.

Fig. 4(a) demonstrates the average optimal EE performance under three systems. The centralized antennas of co-located FD MIMO system are located in the center of the map. It can

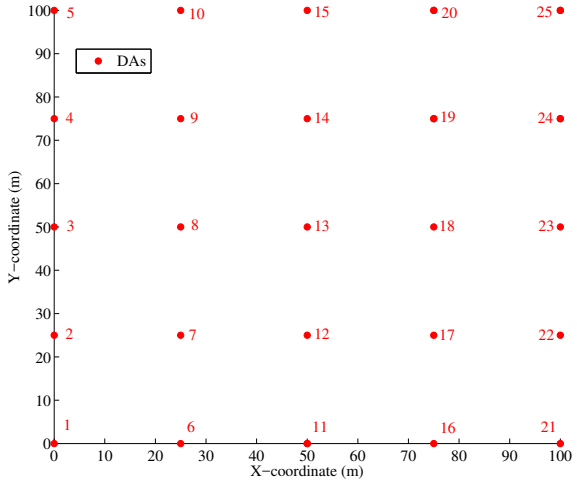

Fig. 2. 25 DAs are uniformly fixed across the area. Without loss of generality, the DAs are indexed by column.

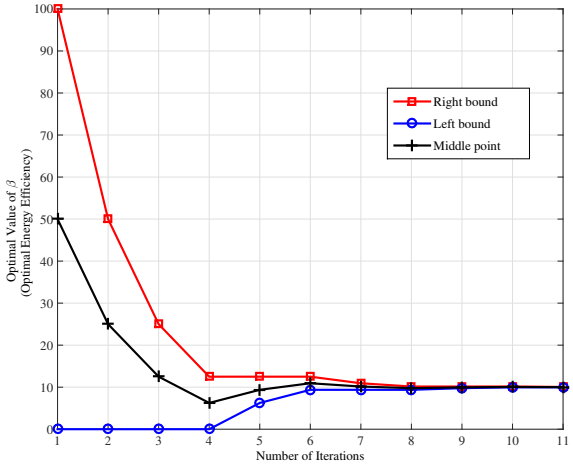

Fig. 3. The convergence behavior on finding optimal β^* (equals to optimal EE) by bisection search, with left bound $\beta_l = 0$, right bound $\beta_r = 100$ and tolerance $\epsilon = 0.1$.

be seen that the proposed system outperforms others in terms of EE at all SIC settings. It is because under the proposed system, the DAs with poor channel condition are deactivated, hence power waste is avoided. Besides, the activated DAs can work in HD modes (only transmit or receive) if the DAs have only reasonable channel gain in one direction. Differently, under the distributed sole-FD system, all activated DAs work in FD mode. This leads to much higher power consumption than the proposed system, but not all DAs can contribute acceptable throughput in bi-direction, hence the EE is degraded. For the co-located FD MIMO system, all the centralized antennas suffer similar PL, leading to poor throughput performance. Besides, high power consumption is incurred due to the fully activated antennas. As a result, the EE performance is the lowest among the three systems. Also, it can be seen that, for all the three systems, EE

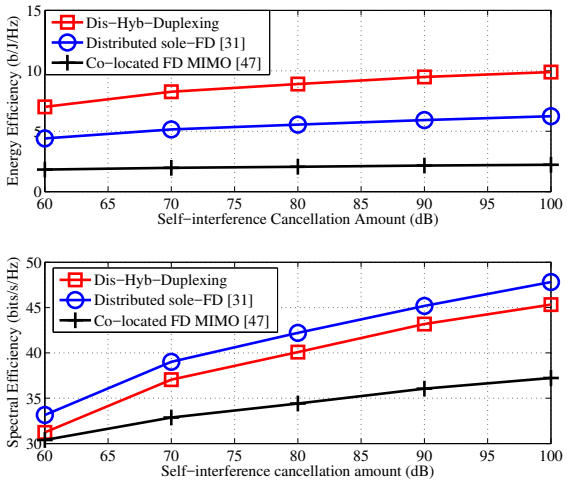

Fig. 4. The effect of SIC amount on EE and SE performances.

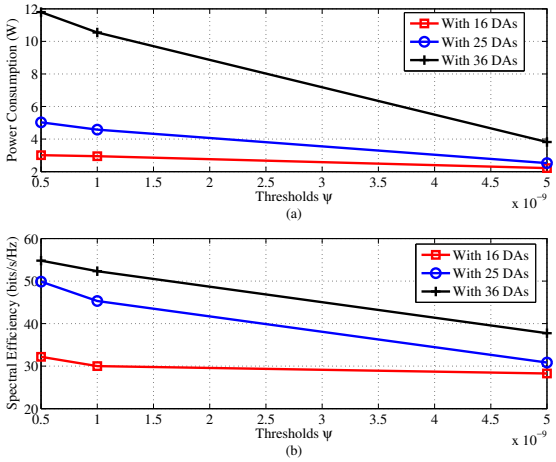

Fig. 6. The effect of DAC threshold value ψ on power consumption and SE performances.

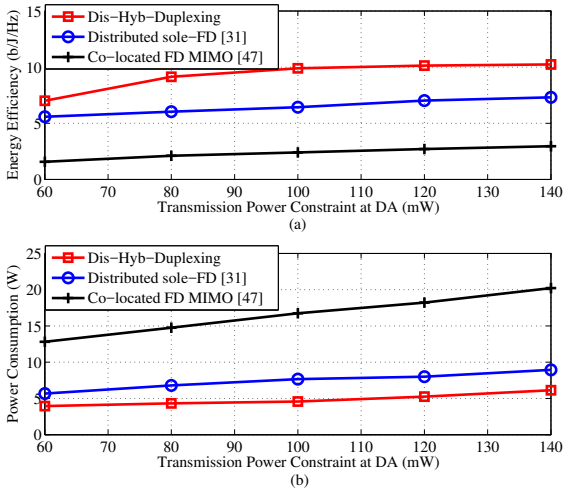

Fig. 5. The effect of power constraint p_{DA} on EE and power consumption.

will be improved with a increased SIC amount due to the less residual self-interference. Fig. 4(b) shows the average SE performance for the three systems. Distributed sole-FD system shows the highest SE performance due to the fact that all antennas are deployed in a distributed manner and work in FD. However, our proposed system has only marginal SE loss compared to the distributed sole-FD, since the DAs far from users are turned off to significantly reduce power consumption. The co-located FD MIMO system demonstrates the poorest SE performance, which is because the co-located antenna deployment is sensitive to the distance between users and antennas and has limited ability in reducing PL. The co-located antenna deployment has also resulted in the strongest self-interference among different antennas. Therefore, the co-located FD MIMO system requires higher SIC amount than the DA deployment. As a result, the co-located FD MIMO system's SE increases more slowly than the DA systems.

Fig. 5(a) and Fig. 5(b) show the effects of power constraint p_{DA} on EE and power consumption, respectively. As shown in Theorem 1, EE is a quasi-concave function with respect to the transmission power. If the power constraint is relatively low, EE is mono-increasing. If the power constraint is relatively high, the global optimal EE can be achieved and the EE performance approaches saturation, as showed in Fig. 5(a). Also, it can be seen that the proposed system outperforms the others with all the transmission power constraints, even when transmission power is only 60 mW. It is because the proposed system can activate/deactivate DAs, and adjust the downlink beamformer and uplink transmission power to make a trade-off between SE and power consumption. Fig. 5(b) shows that the proposed algorithm enables the least power consumption among the three systems. The co-located FD MIMO consumes the highest power. This is because the incurred circuit power consumption in the co-located FD MIMO system plays a dominant role in the total power consumption, and all antennas are kept active to pursue high SE without significant increase in the total power consumption. As a result, the co-located FD MIMO prefers to fully utilize all the transmission power to peruse high SE.

Fig. 6 shows the effect of threshold value ψ on power consumption and SE performances, where only our proposed system is plotted. It can be seen that, less power and lower throughput are demonstrated with a higher threshold, since fewer DAs are activated. This indicates that we can adjust the threshold value to reduce the power consumption, at the cost of SE. On the other hand, a lower threshold allows a better SE performance, while power consumption is relatively higher, which corresponds to a stringent SE QoS scenario. This confirms our analysis in the DA clustering algorithm that by changing the threshold value ψ, the number of activated DAs can be well controlled towards different QoS requirements.

Fig. 7 shows the required number of self-interference reference chains by the three systems. Since the analog and digital domain cancellations need to tap reference signal from every transmit chain and feed it to all receive chains for SIC operation [1] [38], the required number of self-interference

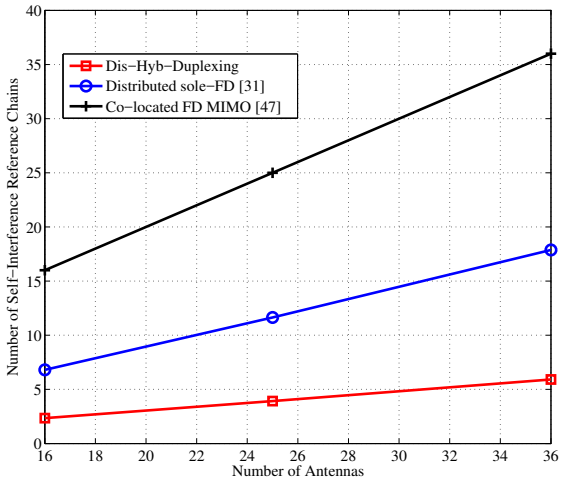

Fig. 7. The effect of number of antennas on the required numbers of reference chains for SIC operation.

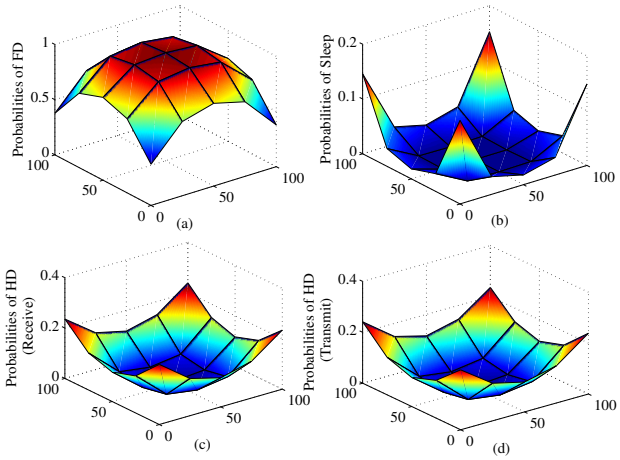

Fig. 8. The average probabilities of DAs' working modes by the DA clustering algorithm with 25 DAs and 16 users, where users' locations are i.i.d and follow uniform distribution.

reference chains has a significant impact on the SIC circuit design. It can be seen that, the proposed system requires the fewest self-interference reference chains. It is because DAs may be in sleep or HD mode, where SIC operation is not needed. Besides, benefiting from the distributed deployment, passive suppression in the propagation domain guarantees a reasonable SIC amount across different DAs, and thus the analog and digital domain SIC are not required across different DAs. For the DA sole-FD systems in [31], DAs can only work in either FD or sleep mode, so the number of reference chains is equivalent to the number of active DAs. For FD MIMO, since all antennas are active and work in FD mode [47] [52], the required number of reference chains is proportional to the number of antennas. As a result, the co-located FD MIMO system requires much more reference chains than others.

Fig. 8 shows the average probabilities of working modes of different DAs, where axes x and y represent distance as

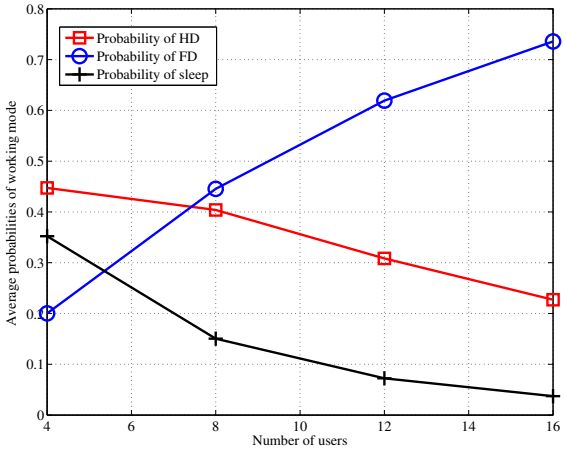

Fig. 9. The effect of the number of users on the probabilities of working modes averaged across 25 DAs.

illustrated in Fig. 2. According to the DA clustering algorithm, only those DAs with good channel conditions in uplink or downlink are turned on for FD mode, and therefore the DAs in the central part of the area have a higher probability of working in FD mode. In contrast, the DAs located at cell edge have a higher probability of working in HD mode or sleeping.

Fig. 9 shows the effect of the number of users on the probabilities of working modes averaged across 25 DAs. With more users, the DAs have a higher probability of working in FD mode while the probabilities of HD and sleep modes degrade. According to the DA clustering algorithm, only the DAs with acceptable channel gain in two directions can work in FD mode. As a result, with more users, the distance between DAs and users is statistically shortened, enabling the increase in the probability of working in FD mode and decrease in the probabilities of HD and sleep modes.

VI. Conclusion

We have proposed a novel bi-directional DA system with hybrid duplexing, which features high EE, low power consumption, and simple SIC circuit design. Based on the proposed system, the EE-oriented resource allocation has been investigated by jointly designing activation/deactivation of transmit/receive chains, downlink beamformer and uplink transmission power. Simulation results show that the proposed design demonstrates significant EE enhancement and power saving over the co-located FD MIMO system and the DA sole-FD system, achieving a reasonable balance between low power consumption and high SE. Moreover, since DAs can flexibly adjust their duplexing and working modes due to the enhanced degree of freedom, the proposed system obtains a much simpler SIC circuit design compared to the two benchmark systems in [31] and [47].

Appendix A
Proof of Theorem 1

Define the superlevel set of $\omega(\boldsymbol{W},\boldsymbol{p})$ as $S_\delta = \{\boldsymbol{W},\boldsymbol{p} \in \boldsymbol{\Theta} | \omega(\boldsymbol{W},\boldsymbol{p}) \geq \delta\}$. According to [19], $\omega(\boldsymbol{W},\boldsymbol{p})$ is jointly quasi-

concave with respect to variables $\boldsymbol{W}$ and $\boldsymbol{p}$ if S_δ is convex for any real number δ. When $\delta \leq 0$, there is no physical meaning. When $\delta \geq 0$, S_δ is equivalent to $S_\delta = \{\boldsymbol{W}, \boldsymbol{p} \in \boldsymbol{\Theta} | p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) - \delta\Big(f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p})\Big) \leq 0\}$. According to our analysis, $P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p})$ is affine with respect to the variables, while $-\delta\Big(f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p})\Big)$ is strictly jointly convex with respect to the variables. Therefore, the summation is strictly convex with respect to the variables, and $\omega(\boldsymbol{W}, \boldsymbol{p})$ is quasi-concave with respect to the variables.

Appendix B
Equivalence between Maximizing Problem $P4$ and Finding the Root of Problem $P5$

The proof is similar to and based on the Lemma 1 in [48]. Suppose that $(\boldsymbol{W^*}, \boldsymbol{p^*})$ is the optimal solution of (23), with the optimal β^*. Define β_0 as the root of $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) = 0$. If $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta_0 p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) = 0$ is obtained at $(\boldsymbol{W_0}, \boldsymbol{p_0})$, we know that there is at least one solution $(\boldsymbol{W_0}, \boldsymbol{p_0})$ satisfying $f_1(\boldsymbol{W_0}, \boldsymbol{p_0}) - f_2^{(n)}(\boldsymbol{W_0}, \boldsymbol{p_0}) - \beta_0 p_{\text{total}}(\boldsymbol{W_0}, \boldsymbol{p_0}) = 0$. Also, we have

$$\beta^* \geq \frac{f_1(\boldsymbol{W_0}, \boldsymbol{p_0}) - f_2^{(n)}(\boldsymbol{W_0}, \boldsymbol{p_0})}{p_{\text{total}}(\boldsymbol{W_0}, \boldsymbol{p_0})} = \beta_0. \tag{24}$$

Due to the optimality, $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta_0 p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) = 0$ means $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta_0 p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) \leq 0$ for all $\boldsymbol{W}, \boldsymbol{p} \in \{\boldsymbol{\Theta}\}$, including $\boldsymbol{W^*}, \boldsymbol{p^*}$. Therefore, we have $\beta^* \leq \beta_0$. As a result, $\beta_* = \beta_0$ is lead.

On the other hand, if $\beta = \beta^*$, at least we have $f_1(\boldsymbol{W^*}, \boldsymbol{p^*}) - f_2^{(n)}(\boldsymbol{W^*}, \boldsymbol{p^*}) - \beta p_{\text{total}}(\boldsymbol{W^*}, \boldsymbol{p^*}) = 0$ and thus $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta^* p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) \geq 0$. Also, because β^* is optimal from (24), for all feasible $(\boldsymbol{W}, \boldsymbol{p}) \in \{\boldsymbol{\Theta}\}$, $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta^* p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) \leq 0$ holds. As a result, the equality $\beta = \beta^*$ yields $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta^* p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) = 0$. In conclusion, the sufficient and necessary condition of $\beta = \beta^*$ is $f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta p_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) = 0$, and the equivalence between maximizing problem $P4$ and finding the root of $P5$ is proven.

Appendix C
Proof of Rank One

The SDP relaxed problem in (23) satisfies the Slater's constraint qualification. To obtain the dual problem, the Lagrange function of the primal problem is given by

$$\begin{aligned} L = {} & f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) + \\ & \sum_{l=1}^{L} \zeta_l \Big(p_{DA} - \sum_{k=1}^{K} Tr(\boldsymbol{W_k} \boldsymbol{F_l})\Big) + \sum_{k=1}^{K} \boldsymbol{W_k} \boldsymbol{Z_k} + \\ & \sum_{u=1}^{U} \varsigma_u (p_{u,max} - p_u) + \sum_{u=1}^{U} \lambda_u p_u, \end{aligned} \tag{25}$$

where λ_u, ζ_l, $\boldsymbol{Z_k}$ and ς_u, for $\forall l \in L, \forall u \in U, \forall k \in K$, are Lagrange multipliers associated with the constraints in (23). According to (25), the Karush-Kuhn-Tucker (KKT) conditions used for the proof include the dual constraints: $\boldsymbol{Z_k^*} \succeq \boldsymbol{0}$ and $\zeta_l^* \geq 0$; complementary slackness: $\boldsymbol{W_k^* Z_k^*} = \boldsymbol{0}$; and the gradient of Lagrange function (25) with respect to $\boldsymbol{W_k}$ vanishing to 0: $\frac{\partial L}{\partial \boldsymbol{W_k}} | \boldsymbol{W_k^*} = 0$. In particular, $\frac{\partial L}{\partial \boldsymbol{W_k}} | \boldsymbol{W_k^*} = 0$ can be further detailed by (26)

$$\boldsymbol{Z_k^*} = \boldsymbol{A_k^*} - \frac{K \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m^T} (\boldsymbol{F_n^H})^T \boldsymbol{H_k}}{\Gamma^*}, \tag{26}$$

where variables $\boldsymbol{A_k^*}$, Γ^* and Λ^* involved in (26) are given by (27) (28) and (29), respectively. By pre-multiplying equality (26) by $\boldsymbol{W_k^*}$, we have

$$\boldsymbol{W_k^* A_k^*} = \boldsymbol{W_k^*} \frac{K \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m^T} (\boldsymbol{F_n^H})^T \boldsymbol{H_k}}{\Gamma}. \tag{30}$$

Since all channel variables in the system are statistically independent and the matrices span the whole signal space, we know that $\boldsymbol{A_k^*}$ is a full-rank matrix [34]. According to the properties of matrix calculus, we have

$$\begin{aligned} & \text{Rank}(\boldsymbol{W_k^*}) = \text{Rank}(\boldsymbol{W_k^* A_k^*}) \\ & = \text{Rank}\Big(\boldsymbol{W_k^*} \frac{K \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m^T} (\boldsymbol{F_n^H})^T \boldsymbol{H_k}}{\Gamma}\Big) \leq \min \\ & \Big\{\text{Rank}(\boldsymbol{W_k^*}), \text{Rank}(\frac{K \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m^T} (\boldsymbol{F_n^H})^T \boldsymbol{H_k}}{\Gamma})\Big\} \\ & \leq \text{Rank}\Big(\frac{K \sum_{m=1}^{L} \sum_{n=1}^{L} t_m t_n \boldsymbol{F_m^T} (\boldsymbol{F_n^H})^T \boldsymbol{H_k}}{\Gamma}\Big) \leq 1. \end{aligned} \tag{31}$$

Now, we have the inequality of $\text{Rank}(\boldsymbol{W_k^*}) \leq 1$. We note that only a null matrix has rank 0 while $\boldsymbol{W_k^*} \neq \boldsymbol{0}$. Thus, $\text{Rank}(\boldsymbol{W_k^*}) = 1$ is obtained.

Appendix D
Proof of Convergence in Inner Layer by the FW Method

Hereby we prove the convergence in the inner layer by the FW method [11]. Let $\Upsilon^{(n+1)}$ denote the optimal value of $(P5)$ at the $(n+1)$-th iteration. According to the updating method in FW, we have

$$\begin{aligned} \Upsilon^{(n+1)} = {} & f_1(\boldsymbol{W^{(n+1)}}, \boldsymbol{p^{(n+1)}}) - f_2^{(n)}(\boldsymbol{W^{(n+1)}}, \boldsymbol{p^{(n+1)}}) \\ & - \beta P_{\text{total}}(\boldsymbol{W^{(n+1)}}, \boldsymbol{p^{(n+1)}}) \\ = {} & \underset{\boldsymbol{W}, \boldsymbol{p} \in \boldsymbol{\Theta}.}{\text{argmax}} f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}). \end{aligned} \tag{32}$$

$\Upsilon^{(n+1)}$ is the optimal value at the $(n+1)$-th iteration, which is greater than any other solutions in the feasible domain. Therefore, we have the following inequality that

$$\begin{aligned} & \underset{\boldsymbol{W}, \boldsymbol{p} \in \boldsymbol{\Theta}.}{\text{argmax}} f_1(\boldsymbol{W}, \boldsymbol{p}) - f_2^{(n)}(\boldsymbol{W}, \boldsymbol{p}) - \beta P_{\text{total}}(\boldsymbol{W}, \boldsymbol{p}) \geq \\ & f_1(\boldsymbol{W^{(n)}}, \boldsymbol{p^{(n)}}) - f_2^{(n)}(\boldsymbol{W^{(n)}}, \boldsymbol{p^{(n)}}) - \beta P_{\text{total}}(\boldsymbol{W^{(n)}}, \boldsymbol{p^{(n)}}) \end{aligned} \tag{33}$$

$$\boldsymbol{A}_{\boldsymbol{k}}^{*}=\sum_{k'\neq K}^{K}\frac{\sum_{m=1}^{L}\sum_{n=1}^{L}t_m t_n \boldsymbol{F}_{\boldsymbol{m}}^{T}(\boldsymbol{F}_{\boldsymbol{n}}^{H})^{T}\ \boldsymbol{H}_{\boldsymbol{k'}}}{a_{k'}^{(n)}}+\frac{\beta}{\eta}\boldsymbol{I}+\sum_{l=1}^{L}\zeta_l^{*}\boldsymbol{F}_{\boldsymbol{l}}^{T}+$$
$$\sum_{u=1}^{U}\frac{\frac{1}{\alpha}\mathrm{diag}\Big(\boldsymbol{H}_{\boldsymbol{SI}}(\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}t_m t_n \boldsymbol{F}_{\boldsymbol{m}}^{T}\boldsymbol{F}_{\boldsymbol{n}}^{H})^{T}\boldsymbol{H}_{\boldsymbol{SI}}^{H}\Big)(\sum_{i=1}^{L}\sum_{j=1}^{L}r_i r_j \boldsymbol{F}_{\boldsymbol{i}}\boldsymbol{G}_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{j}}^{H})}{b_u^{(n)}} \tag{27}$$
$$-\frac{U\frac{1}{\alpha}\mathrm{Tr}\Big(\mathrm{diag}\Big(\boldsymbol{H}_{\boldsymbol{SI}}(\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}t_m t_n \boldsymbol{F}_{\boldsymbol{m}}^{T}\boldsymbol{F}_{\boldsymbol{n}}^{H})^{T}\boldsymbol{H}_{\boldsymbol{SI}}^{H}\Big)(\sum_{i=1}^{L}\sum_{j=1}^{L}r_i r_j \boldsymbol{F}_{\boldsymbol{i}}\boldsymbol{G}_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{j}}^{H})\Big)}{\Lambda^{*}}.$$

$$\Gamma^{*}=\sum_{k=1}^{K}\log_2\Big(\sum_{k'=1}^{K}\mathrm{Tr}(\sum_{m=1}^{L}\sum_{n=1}^{L}t_m t_n \boldsymbol{F}_{\boldsymbol{m}}\boldsymbol{W}_{\boldsymbol{k'}}^{*}\boldsymbol{F}_{\boldsymbol{n}}^{H}\boldsymbol{H}_{\boldsymbol{k}})+\sum_{u=1}^{U}p_u^{*}|e_{u,k}|^2+\sigma_k^2\Big). \tag{28}$$

$$\Lambda^{*}=\sum_{u=1}^{U}\log_2\Bigg(\sum_{u'\neq u}^{U}p_{u'}^{*}\mathrm{Tr}(\boldsymbol{G}_{\boldsymbol{u'}}\sum_{i=1}^{L}\sum_{j=1}^{L}r_i r_j \boldsymbol{F}_{\boldsymbol{i}}\boldsymbol{G}_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{j}}^{H})+\sigma^2\mathrm{Tr}(\sum_{i=1}^{L}\sum_{j=1}^{L}r_i r_j \boldsymbol{F}_{\boldsymbol{i}}\boldsymbol{G}_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{j}}^{H})+$$
$$\frac{1}{\alpha}\mathrm{Tr}\Big(\mathrm{diag}\Big(\boldsymbol{H}_{\boldsymbol{SI}}(\sum_{k=1}^{K}\sum_{m=1}^{L}\sum_{n=1}^{L}t_m t_n \boldsymbol{F}_{\boldsymbol{m}}\boldsymbol{W}_{\boldsymbol{k}}^{*}\boldsymbol{F}_{\boldsymbol{n}}^{H})\boldsymbol{H}_{\boldsymbol{SI}}^{H}\Big)(\sum_{i=1}^{L}\sum_{j=1}^{L}r_i r_j \boldsymbol{F}_{\boldsymbol{i}}\boldsymbol{G}_{\boldsymbol{u}}\boldsymbol{F}_{\boldsymbol{j}}^{H})\Big)\Bigg). \tag{29}$$

Since we have the equalities that $f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}}) = f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$ and $\nabla f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}}) = \nabla f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$, (33) can be further derived as

$$\begin{aligned}&f_1(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-\beta P_{\mathrm{total}}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})\\&=f_1(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-\beta P_{\mathrm{total}}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})\\&\geq f_1(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-f_2^{(n-1)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})\\&\quad-\beta P_{\mathrm{total}}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})=\Upsilon^{(n)}.\end{aligned} \tag{34}$$

As a result, $\Upsilon^{(n+1)} \geq \Upsilon^{(n)}$ is led, meaning the value of (23) is non-decreasing after each iteration. Also, the total throughput is upper bounded by the transmission power constraints $(\tilde{C}1)$ and $(C2)$, and thus the FW method confirms the convergence. Then we prove that the convergence leads to a KKT point. Since the Lagrange function of (23) has been given by (25), all the KKT conditions of the optimal value at iteration n are given by equations (35)-(40)

$$\begin{aligned}&\nabla_{\boldsymbol{W_k}}f_1(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-\nabla_{\boldsymbol{W_k}}f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})+\boldsymbol{Z_k}\\&-\nabla_{\boldsymbol{W_k}}\beta P_{\mathrm{total}}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-\sum_{l=1}^{L}\zeta_l\boldsymbol{F_l}=0,\end{aligned} \tag{35}$$

$$\begin{aligned}&\nabla_{p_u}f_1(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})-\nabla_{p_u}f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})\\&-\varsigma_u+\lambda_u-\nabla_{p_u}\beta P_{\mathrm{total}}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})=0,\end{aligned} \tag{36}$$

$$\zeta_l\Big(p_{DA}-\sum_{k=1}^{K}Tr(\boldsymbol{W_k^{(n)}}\boldsymbol{F_l})\Big)=0, \tag{37}$$

$$Tr(\boldsymbol{W_k^{(n)}}Z_k)=0 \tag{38}$$

$$\varsigma_u(p_{u,max}-p_u^{(n)})=0, \tag{39}$$

$$\lambda_u p_u^{(n)}=0, \tag{40}$$

for $\forall k \in K$, $\forall u \in U$, and $\forall l \in L$. According to the equality that $\nabla f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}}) = \nabla f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$, we can replace $\nabla_{\boldsymbol{W_k}}f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$ and $\nabla_{p_u}f_2^{(n)}(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$ by $\nabla_{\boldsymbol{W_k}}f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$ and $\nabla_{p_u}f_2(\boldsymbol{W^{(n)}},\boldsymbol{p^{(n)}})$, which are exactly the KKT conditions of the original problem $f_1(\boldsymbol{W},\boldsymbol{p}) - f_2(\boldsymbol{W},\boldsymbol{p}) - P_{\mathrm{total}}(\boldsymbol{W},\boldsymbol{p})$. As a result, the FW method leads to a convergence in the inner layer.

REFERENCES

[1] M. Duarte, A. Sabharwal, V. Aggarwal, R. Jana, and K. Ramakrishnan, "Design and characterization of a full-duplex multi-antenna system for WiFi networks," *IEEE Trans. Veh. Tech.*, vol. 63, no. 3, pp. 1160-1177, Mar. 2014.

[2] X. Zhu and R. D. Mruch, "Layered space-frequency equalization in a single-carrier MIMO system for frequency-selective channels," *IEEE Trans. Wireless Commun.*, vol. 3, no. 3, pp. 701-708, May 2004.

[3] H. Xie, F. Gao, S. Zhang, and S. Jin, "A unified transmission strategy for TDD/FDD massive MIMO systems with spatial basis expansion model," *IEEE Trans. Veh. Tech.*, vol. 66, no. 4, pp. 4170-3184, Mar. 2017.

[4] Z. Wei, X. Zhu, S. Sun, and Y. Huang, "Energy efficiency oriented cross-Layer resource allocation for multiuser full-duplex decode-and-forward indoor relay systems at 60 GHz," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 12, pp. 3366-3379, Dec. 2016.

[5] D. Kim, H. Lee, and D. Hong, "A survey of in-band full-duplex transmission: from the perspective of PHY and MAC layers," *IEEE Commun. Surveys Tut.*, vol. 17, no. 4, pp. 2017- 2046, Jul. 2015.

[6] E. Everett, A, Sahai, and A. Sabharwal, "Passive self-interference suppression for full-duplex infrastructure nodes," *IEEE Trans. Wireless Commun.*, vol. 13, no. 2, pp. 680-694, Feb. 2014.

[7] B. Radunovic, D. Gunawardena, P. Key, A. Proutiere, N. Singh, V. Balan, and G. Dejean, "Rethinking indoor wireless mesh design: low power, low frequency, full-duplex," Microsoft Technical Report, Tech. Rep., 2009.

[8] A. Almradi and K. A. Hamdi, "On the outage probability of MIMO full-duplex relaying: impact of antenna correlation and imperfect CSI," *IEEE Trans. Veh. Tech.*, vol. 66, no. 5, pp. 3957-3965, May 2017.

[9] L. Song, Y. Li, and Z. Han, "Resource allocation in full-duplex communications for future wireless networks," *IEEE. Wireless Commun.*, vol. 22, no. 4, pp. 88-96, Aug. 2015.

[10] P. Day, A. R. Margetts, D. W. Bliss, and P. Schniter, "Full-duplex bidirectional MIMO: achievable rates under limited dynamic range," *IEEE Trans. Sig. Process.*, vol. 6, no. 7, pp. 3702-3713, Nov. 2003.

[11] D. Nguyen, L. Tran, P. Pirinen, and M. Latva-aho, "On the spectral efficiency of full-duplex small cell wireless system," *IEEE Trans. Wireless Commun.*, vol. 13, no. 9, pp. 4896-4910, Sep. 2014.

[12] S. Huberman and T. L. Ngoc, "MIMO full-duplex precoding: a joint beamforming and self-interference cancellation structure," *IEEE Trans. Wireless Commun.*, vol. 14, no. 4, pp. 2217-2205, Apr. 2015.

[13] S. Akbar, Y. Deng, A. Nallanathan, M. Elkashlan and G.K Karagiannidis, "Massive multiuser MIMO in heterogeneous cellular networks with full duplex small cells," *to appear in IEEE Trans Wireless Commun.*, DOI 10.1109/TCOMM.2017.2728536.

[14] H. Tabassum, A. H. Sakr, and E. Hossain, "Analysis of massive MIMO-enabled downlink wireless backhauling for full duplex small cells," *IEEE Trans Wireless Commun.*, vol. 64, no. 6, pp. 2354-2369, Jun. 2016.

[15] H. Cui, M. Ma, L. Song, and B. Jiao, "Relay selection for two-way full duplex relay networks with amplify-and-forward protocol," *IEEE Trans. on Wireless Commun.*, vol. 13, no. 7, pp. 3768-3777, May 2014.

[16] S. Li, M. Zhou, J. Wu, L. Song, Y. Li, and H. Li, "X-duplex relaying: Adaptive antenna configuration," *IEEE. Commun. Lett.*, vol. 21, no. 5, pp. 1083-1086, May 2017.

[17] M. Jain *et al.,* "Practical, real-time, full duplex wireless," in *Proc. ACM MobiCom'10*, New York, USA, Sep. 2010, pp. 301-312.

[18] Z. Wei, X. Zhu, S. Sun, Y. Huang, L. Dong, and Y. Jiang, "Full-duplex vs. half-duplex amplify-and-forward relaying: which is more energy efficient in 60 GHz dual-hop indoor wireless systems?" *IEEE J. Sel. Areas Commun.*, vol. 33, no. 12, pp. 2936-2947, Dec. 2015.

[19] C. Xiong, G. Y. Li, S. Zhang, Y. Chen, and S. Xu, "Energy-and spectral-efficiency tradeoff in downlink OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 10, no. 11, pp. 3874-3886, Nov. 2011.

[20] C. Xiong, G. Y. Li, Y. Liu, Y. Chen, and S. Xu, " Energy-efficient design for downlink OFDMA with delay-sensitive traffic," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 3085-3095, Jun. 2013.

[21] D. W. K. Ng, E. S. Lo, and R. Schober, "Energy-efficient resource allocation in OFDMA systems with large number of base station antennas," *IEEE Trans. Wireless Commun.*, vol. 11, no. 9, pp. 3292-3304, Sep. 2012.

[22] X. Gao, O. Edfors, F. Tufvesson, and E. G. Larsson, "Massive MIMO in real propagation environments: do all antennas contribute equally?" *IEEE Trans. Commun.*, vol. 63, no. 11, pp. 3917-3928, Nov. 2015.

[23] J. Joung, Y. K. Chia, and S. Sun, "Energy-efficient, large-scale distributed-antenna system (L-DAS) for multiple users," *IEEE Trans. Sig. Process.*, vol. 8, no. 5, pp. 964-965, Oct. 2014.

[24] FCC Order 11-50 (PDF). FCC. Retrieved 2014-09-25.

[25] J. Wang and L. Dai, "Downlink rate analysis for virtual-cell based large-scale distributed antenna systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1998-2011, Mar. 2016.

[26] C. Loyez, M. Bocquet, C. Lethien, and N. Rolland, "A distributed antenna system for indoor accurate WiFi localization," *IEEE Antennas and Wireless Propag. Lett.*, VOL. 14, pp. 1184-1187, Jan. 2015.

[27] H. Wang, H. Hou, and J. Chen, "Design and analysis of an antenna control mechanism for time division duplexing distributed antenna systems over high-speed rail communications," *IEEE Trans. Emerg. Topics Comput.*, vol. 4, no. 4, pp. 516-527, Dec. 2016.

[28] J. Wang, H. Zhu, and N. J. Gomes, "Distributed antenna systems for mobile communications in high speed trains," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 4, pp. 675-683, May. 2012.

[29] C. He, G. Y. Li, F. Zheng, and X. You, "Power allocation criteria for distributed antenna systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5083-5090, Nov. 2015.

[30] X. Ge, H. Jin, and V. C. M. Leung, "Opportunistic downlink scheduling with resource-based fairness and feedback reduction in distributed antenna systems," *IEEE Trans. Veh. Technol.*, vol. 64, no. 11, pp. 5083-5090, Nov. 2015.

[31] D. W. K. Ng, Y. Wu, and R. Schober, "Power efficient resource allocation for full-duplex radio distributed antenna networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2896-2911, Apr. 2016.

[32] R. Bolla, R. Bruschi, F. Davoli, and F. Cucchietti, "Energy efficiency in the future Internet: a survey of existing approaches and trends in energy-aware fixed network infrastructures," *IEEE Commun. Surveys Tut.*, vol. 13, no. 2, pp. 223-244, Jul. 2011.

[33] G. Liu, F. R. Yu, H. Ji, and V. C. M. Leung, "Energy-efficient resource allocation in cellular networks with shared full-duplex relaying," *IEEE Trans. Veh. Technol.*, vol. 64, no. 8, pp. 3711-3724, Aug. 2015.

[34] Y. Sun , D. W. K. Ng, J. Zhu, and R. Schober, "Multi-objective optimization for robust power efficient and secure full-duplex wireless communication systems," *IEEE Trans. Wireless Commun.*, vol. 15, no. 8, pp. 5511-5526, Aug. 2016.

[35] A. C. Cirik, O. Taghizadeh, R. Mathar, and T. Ratharajah, "QoS considerations for full-duplex multiuser MIMO systems," *IEEE Wireless Commun. Lett.*, vol. 5, no. 1, pp. 36-39, Feb. 2016.

[36] C. Nam, C. Joo, and S. Bahk, "Joint subcarrier assignment and power allocation in full-duplex OFDMA networks," *IEEE Trans. Wireless Commun.*, vol. 14, no. 6, pp. 3109-3119, Jun. 2015.

[37] G. Liu, F. R. Yu, H. Ji, V. C. M. Leung, and X. Li, "In-band full-duplex relaying: a survey, research issues and challenges," *IEEE Commun. Surveys Tuts.*, 2015. 500-523, Jan. 2015.

[38] D. Bharadia, E. McMilin, and S. Katti, "Full duplex radios," *in Proc. ACM SIGCOMM'13*, Hong Kong, China, Aug. 2013, pp. 375-386.

[39] H. Ren, N. Liu, C. Pan, L. Hanzo, "Joint fronthaul link selection and transmit precoding for energy efficiency maximization of multiuser MIMO-aided distributed antenna systems," *to appear in IEEE Trans. Wireless Commun.*, DOI:10.1109/TCOMM.2017.2728526.

[40] H. Kim, S. Lee, C. Song, K. Lee, and I. Lee, "Optimal power allocation scheme for energy efficiency maximization in distributed antenna systems," *IEEE Trans. Commun.*, vol. 63, no. 2, pp. 431-440, Dec. 2015.

[41] M. Tham, S. Chien, D. W. Holtby, and S. Alimov, "Energy-efficient power allocation for distributed antenna systems with proportional fairness," *Green Commun. and Net*., vol. 1, no. 2, pp. 145-157, Apr. 2017.

[42] S. Cui, A. J. Goldsmith, and A. Bahai, "Energy-constrained modulation optimization," *IEEE Trans. Wireless Commun.*, vol. 4, no. 5, pp. 2349-2360, Sep. 2005.

[43] EARTH Project. D2.3 Energy efficiency analysis of the reference systems, areas of improvements and target breakdown. [Online]. Available: http://www.ict-earth.eu/publications/publications.html

[44] L. Lu, G. Y. Li, A. L. Swindlehurst, A. Ashikhmin, and R. Zhang, "An overview of massive MIMO: benefits and Challenges," *IEEE J. Sel. Topics Sig. Process.*, vol. 8, no. 5, pp. 742-758, Oct. 2014.

[45] V. Syrjala, M. Valkama, L. Anttila, T. Riihonen, and D. Korpi, "Analysis of oscillator phase-noise effects on self-interference cancellation in full-duplex OFDM radio transceivers," *IEEE Trans. Wireless Commun.*, vol. 13, no. 06, pp. 2977-2990, Jun. 2014.

[46] T. Riihonen, S. Werner, and R. Wichman, "Optimized gain control for single-frequency relaying with loop interference," *IEEE Trans. Sig. Process.*, vol. 59, no. 12, pp. 5983-5993, Dec. 2011.

[47] D. Nguyen, L. N. Tran, P. Pirinen, and M. L. Aho, "Precoding for full-duplex multiuser MIMO systems: spectral and energy efficiency maximization," *IEEE Trans. Sig. Process.*, vol. 61, no. 16, pp. 4038-4050, Aug. 2013.

[48] X. Xiao, X. Tao, and J. Lu, "Energy-efficient resource allocation in LTE-based MIMO-OFDMA systems with user rate constraints," *IEEE Trans. Veh. Tech.*, vol. 64, no. 1, pp. 185-197, Jan. 2015.

[49] Y. Sun, D. W. K. Ng, and R. Schober, "Multi-objective optimization for power efficient full-duplex wireless communication systems," in *Proc. IEEE Globecom'15*, San Diego, USA, Dec. 2015, pp. 1-6.

[50] M. Jaggi, "Revisiting Frank-Wolfe projection-free sparse convex optimization" in *Proc. ICML'13*, Atlanta, USA, Jun. 2013, pp. 427-435.

[51] Y. Cheng, M. Pesavento, and A. Philipp, "Joint network optimization and downlink beamforming for CoMP transmissions using mixed integer conic programming," *IEEE Trans. Sig. Process.*, vol. 61, no. 16, pp. 3972-3987, Aug. 2013.

[52] M. Duarte, C. Dick, and A. Sabharwal, "Experiment-Driven Characterization of Full-Duplex Wireless Systems," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 12, pp. 4296-4307, Dec. 2012.

Zhongxiang Wei (S'15-M'17) received the Ph.D. degree in Electrical Engineering and Electronics from the University of Liverpool, Liverpool, U.K, in 2017. From Mar. 2016 to Mar. 2017, he was with the Institution for Infocomm Research (I2R), Singapore, as a Research Assistant. His research interests include green communications, full-duplex, millimeter-wave communications, resource allocation, and algorithm design. He received graduate China National Scholarship Award in 2012 and CSC Outstanding Self-Financed Scholarship in 2017.

Sumei Sun (F'16) is currently the Head of the Communications and Networks Cluster, Institute for Infocomm Research, Agency for Science, Technology, and Research, Singapore, focusing on smart communications and networks for robust, QoS/QoE-guaranteed, and energy- and spectrum-efficient connectivity for human, machine, and things. She publishes actively in IEEE journals and conferences, and collaborates closely in research with the industry. She is a Distinguished Lecturer of the IEEE Vehicular Technology Society 2014-2018, a Distinguished Visiting Fellow of the Royal Academy of Engineering, U.K., in 2014, and Vice Director of IEEE Communications Society Asia Pacific Board since 2016. She has also been actively contributing to organizing IEEE conferences in different roles, including her recent services as Executive Vice Chair of Globecom 2017, Symposium Co-Chair of ICC 2015 and 2016, Track Co-Chair of IEEE VTC 2016 Fall, VTC 2017 Spring, etc. She was an Editor of the IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY (TVT) during 2011-2017, and Editor of the IEEE WIRELESS COMMUNICATION LETTERS during 20112016. She serves as an Area Editor of IEEE TVT Since July 2017, and Editor of the IEEE COMMUNICATIONS SURVEYS AND TUTORIALS, since 2015. She received the Top Associate Editor award in 2011, 2012, and 2015, all from IEEE Transactions on Vehicular Technology.

Xu Zhu (S'02-M'03-SM'12) received the BEng degree (with the first class honors) in Electronics and Information Engineering from Huazhong University of Science and Technology, Wuhan, China, in 1999, and the PhD degree in Electrical and Electronic Engineering from the Hong Kong University of Science and Technology, Hong Kong, in 2003. Since May 2003, she has been with the Department of Electrical Engineering and Electronics, the University of Liverpool, Liverpool, UK, where she is currently a Reader. Dr. Zhu has over 150 peer-reviewed publications on communications and signal processing. She has served as an Editor for the IEEE Transactions on Wireless Communications from 2012-2017, and a Guest Editor for a number of international journals such as Electronics. She has organized various international conferences as Chair, such as Symposium Co-Chair of the IEEE ICC 2016 and 2019, Vice Chair of the 2006 and 2008 ICARN International Workshops, Program Chair of the ICSAI 2012, and Publicity Chair of the IEEE IUCC-2016. Her research interests include MIMO, channel equalization, resource allocation, optimization, cooperative communications, green communications etc..

Yi Huang (S'91-M'96-SM'06) received the DPhil in Communications from the University of Oxford, UK in 1994. He has been conducting research in wireless communications, applied electromagnetics, radar and antennas since 1987. His experience includes 3 years spent with NRIET (China) as a Radar Engineer and various periods with the Universities of Birmingham, Oxford, and Essex at the UK. He worked as a Research Fellow at British Telecom Labs in 1994, and then joined the University of Liverpool, UK in 1995, where he is now a full Professor and Deputy Head of the Department of Electrical Engineering and Electronics. Dr. Huang has published over 300 refereed papers, received many research grants from various funding bodies and acted as consultant for various companies. He has been a keynote/invited speaker and organizer of many conferences and workshops (*e.g.*, WiCom2010, IEEE iWAT 2010 and LAPC2012). He is the Editor-in-Chief of Wireless Engineering and Technology and has been an Editor or Guest Editor for four international journals. He is a UK national representative of EU COST-IC1102 and Executive Committee Member and Fellow of the IET.

Jingjing Wang (S'14) received his B.S. degree in Electronic Information Engineering from Dalian University of Technology in 2014 with the highest honors. He currently works for his PhD degree at Complex Engineered Systems Lab (CESL) in Tsinghua University, Beijing. From Jan. 2016 to Mar. 2016, he visited Wireless Networks and Decision Systems (WNDS) Group, Singapore University of Technology and Design as a visiting student. Since Otc. 2017, he has been visiting the Telecommunications Group, University of Southampton as a joint PhD student. His research interests include the resource allocation and network association, learning theory aided modeling, analysis and signal processing, and information diffusion theory. He received Tsinghua GuangHua Scholarship in 2016 and graduate China National Scholarship Award in 2017.